Corso di Laurea in Ingegneria Informatica
Anno Accademico 2015/2016
Calcolo delle Probabilità e Statistica Matematica

Nome ....

N. Matricola .... Ancona, 8 giugno 2016

1. Un farmaco per la cura di una data malattia viene somministrato al 75 % di una popolazione. Con l'assunzione del farmaco la probabilità di contrarre la malattia è del 10 %, mentre è del 50 % senza il farmaco.

   (a) Qual è l'incidenza della malattia sulla popolazione (cioè la probabilità che una persona scelta a caso si ammali)?

   (b) Se un individuo scelto a caso non ha contratto la malattia, qualè la probabilità che abbia assunto il farmaco?

Soluzione. Definiamo gli eventi $E =$ "la persona ha assunto il framaco $M =$"la persona è malata. Dai dati del problema abbiamo:

$$P(E) = 0.75, \quad P(\bar{E}) = 0.25, \quad P(M|E) = 0.1, \quad P(M|\bar{E}) = 0.5$$

(a) Usando il teorema delle probabilità totali:

$$P(M) = P(M|E)\,P(E) + P(M|\bar{E})\,P(\bar{E}) = 0.1 \times 0.75 + 0.5 \times 0.25 = 0.2$$

L'incidenza della malattia è dunque del 20 %.

(b) Usando la formula di Bayes:

$$P(E|\bar{M}) = \frac{P(\bar{M}|E)\,P(E)}{P(\bar{M})} = \frac{[1 - P(M|E)]\,P(E)}{1 - P(M)} = \frac{0.9 \times 0.75}{0.8} = 0.84$$

2. Siano X e Y due variabili aleatorie non negative la cui distribuzione congiunta è uniforme nel triangolo di vertici $A = (0, 0)$, $B = (1, 0)$ e $C = (1, 1)$.

   (a) Determinare il valore della distribuzione nel dominio;

   (b) determinare le densità marginali $f_X(x)$ e $f_Y(y)$, specificando esplicitamente per entrambe l'intervallo di definizione;

   (c) verificare se X ed Y sono indipendenti;

   (d) determinare media e varianza di X e Y;

   (e) determinare il valor medio del prodotto $X\,Y$;

   (f) determinare il coefficiente di correlazione di X e Y.

Soluzione. Sia R la regione del piano corrispondente al triangolo di vertici A, B e C. Dai dati del problema abbiamo:

$$f_{XY}(x,y) = k, \quad (x,y) \in R, \qquad f_{XY}(x,y) = 0, \quad \text{altrimenti}$$

(a)

$$\int\int_R k\,dx\,dy = 1 \qquad \int_0^1 dx \int_0^x dy\,k = 1 \qquad \frac{1}{2}\,k = 1,$$

da cui $k = 2$;

(b)

$$f_X(x) = \int_0^x f_{XY}(x,y)\,dy = 2\,x \quad \text{per} \quad 0 \le x \le 1$$
$$f_Y(y) = \int_y^1 f_{XY}(x,y)\,dx = 2\,(1-y) \quad \text{per} \quad 0 \le y \le 1;$$

(c) X ed Y non sono indipendenti, perchè $f_{XY}(x,y) \neq f_X(x)\,f_Y(y)$;

(d)

$$E[X] = \int_0^1 x\,f_X(x)\,dx = \int_0^1 2\,x^2\,dx = \frac{2}{3}$$
$$E[Y] = \int_0^1 y\,f_Y(y)\,dy = \int_0^1 2\,y\,(1-y)\,dy = \frac{1}{3}$$
$$E[X^2] = \int_0^1 x^2\,f_X(x)\,dx = \int_0^1 2\,x^3\,dx = \frac{1}{2}$$
$$E[Y^2] = \int_0^1 y^2\,f_Y(y)\,dy = \int_0^1 2\,y^2\,(1-y)\,dy = \frac{1}{6}$$
$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$
$$Var(Y) = E[Y^2] - E[Y]^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$$

(e)

$$E[X\,Y] = \int_R x\,y\,f_{XY}(x,y)\,dx\,dy = \int_0^1 dx \int_0^x dy\,x\,y\,2 =$$
$$= 2\int_0^1 dx\,x \int_0^x dy\,y = 2\int_0^1 dx\,x\,\frac{x^2}{2} = \frac{1}{4};$$

(f)

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{E[X\,Y] - E[X]\,E[Y]}{\sqrt{Var(X)\,Var(Y)}} = \frac{1/4 - 2/9}{1/18} = \frac{1}{2}$$

3. Il reddito medio annuo pro-capite della popolazione italiana è di circa 20 mila euro. In una certa regione, si pensa che il valore sia più alto rispetto alla media nazionale, e quindi si esegue un test su un campione di 10 individui ottenendo i seguenti risultati (in migliaia di euro):

$$32.0,\ 18.9,\ 21.8,\ 23.1,\ 18.0,\ 16.0,\ 26.2,\ 18.4,\ 19.0,\ 25.0$$

Sulla base di questo campione, si può accettare, con un livello di significatività del 5%, l'ipotesi che il reddito annuo pro-capite medio in quella regione sia superiore alla media nazionale?

(Si prenda come ipotesi nulla
H_0 := *la popolazione da cui è estratto il campione ha reddito medio* $\mu = \mu_0 = 20$ *mila euro con varianza incognita.*
e come ipotesi alternativa
H_1 := *la media è superiore a* μ_0*, ovvero* $\mu > \mu_0$*.)*

Soluzione. Il campione ha media campionaria $\overline{X}_n = 21.8$, varianza campionaria $S_n^2 = 23.5$ e scarto quadratico medio $S_n = 4.8$ (dopo arrotondamento ad una cifra significativa). La variabile standardizzata è

$$Z = \frac{\overline{X}_n - \mu_0}{S_n}\sqrt{n}$$

e la regione di rigetto Z è data da $Z > t_{0.95}(9) = 1.833$. Dai dati abbiamo

$$Z = \frac{21.8 - 20}{4.8}\sqrt{10} = 1.18$$

e quindi il campione non permette di regettare l'ipotesi nulla.